

A New Anticipatory Protocol for Removing of Fraud in Online Through Probit Auction Model

M.Priyanka^{#1}, G.Seetharatnam^{*2}, Kunjam Nageswara Rao^{*3},
Peri.Srinivasa Rao^{*4}

Department of Computer Science & Systems Engineering,
Andhra University College of Engineering,
Visakhapatnam District, AP, India.

Abstract

Now a day's e-commerce is increasing its reputation by making a lot of supply and demand for the end users all around the world. It is expected very fast in its popularity than predicted as it is up over 500% compared in the past. This was increasing the user attention mainly due to the ease of customers to do shopping in online rather than without spending much time by shopping manually. As this e-commerce gives a lot of burden free product for the customers, there are also some criminals who try to do fraud and get profit in illegal ways. As people are enjoying the advantages from online trading, traitors are also taking advantages to accomplish betrayal activities against candid parties to obtain dishonest profit. In this paper in order to detect and prevent such illegal and betrayal activities online probit fraud-detection moderation systems by using machine learned models are commonly applied in practice. In this paper, we show that this model can probably distinguish more deceptions and extensively decrease customer complaints which are based on a real-world online auction fraud detection data compared to several baseline models and the human-tuned rule-based system.

Keywords

Online Shopping, Fraud Detection, Online Probit Model, Online Auction, Machine Learned Models, Betrayal, Moderation, World Wide Web (WWW).

1. Introduction

Now a day's e-commerce is increasing its reputation by making a lot of supply and demand for the end users all around the world. It is expected very fast in its popularity than predicted as it is up over 500% compared in the past. As various websites are allowing Internet users to purchase and trade products which profits everyone in terms of effortlessness and effectiveness [1], [2]. People who shop are enjoying the advantages from online trading; at the same time traitors are also taking advantages to accomplish deceptive activities against candid parties to obtain dishonest profit. Since the emergence of World Wide Web (WWW) as in [3], electronic commerce, which is commonly called as e-commerce as in [4] become more and more popular in recent years.

Now a day's users don't think of taking a stroll through the super market for purchasing any products in the shopping mail, but he tries to go

buying the same item through online in which he is not asked to carry a stroll for carrying items, just he sits in front of his system and purchase the items whatever he needed. The scenario is not limited to a single item alone; it even covers a wide range of products like home appliances, consumer electronic goods, books, apparels, travelling packages, gift articles, toys, dresses etc. and even the electronic content itself. By using this e-Commerce sites as mentioned in [4] then, you can buy almost anything you wish for without actually touching the product physically and inquiring the salesman for a number of times before placing the final order. In existing online shopping business model sellers as in [5] sell their products or services at preset price, where buyers can choose what product best suites them which is of good deal. Online Auction system however is a different business model where the items are sold through price bidding done by various users. To provide some assurance against fraud and to give confidence to online auction services as in [6] E-commerce sites provide insurance to victims for those who loss up to a certain amount.

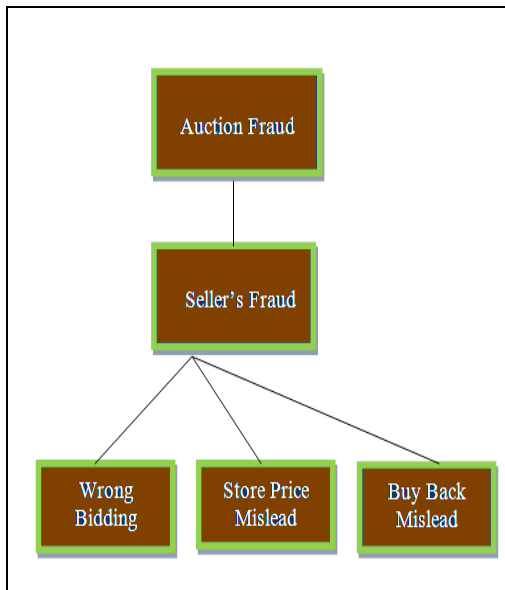


Figure. 1. Represents Online Auction Fraud

In recent days for purchasing an item in online auction website they are to be validated with

e-mail, SMS, or phone call verifications as in [7]. In this paper, we mainly study the application of a proactive moderation system as in [8] for fraud detection, where hundreds and thousands of new auction cases are created every day. Due to the limited expert resources only 20%-40% of cases can be reviewed and labeled but not all. Therefore, it is necessary to develop pre-screening moderation system as in [8] that only directs suspicious/illegal cases for expert review and passes the rest as clean cases. Human experts are also willing to test and see the results of online feature selection to monitor the effectiveness and stability as in [9] of the current set of features, so that they can understand the pattern of frauds done by fraudulent sellers and further add or remove some features as shown in Figure. 1. In this paper we study the problem of building online modeling system for the auction fraud detection moderation system as in [10].

2. Related Work

The main important issue in the online is fraud/fault that is caused in online shopping. There are many articles on websites which teach people how to avoid online auction fraud categorizes auction fraud into several types and proposes strategies to fight them. Some Reputation systems are used extensively by web-sites to detect auction frauds, although many of them use naive approaches. Summarized several key properties of a good reputation system and also the challenges for the modern reputation systems to elicit user feedback. Other representative work connecting reputation systems with online auction fraud detection include where the last work introduced a Markov random field model with a belief propagation algorithm for the user reputation.

In this paper we treat the fraud detection problem as a binary classification problem which has two possibilities. The most frequently used models for binary classification include logistic regression, probit regression, support vector machine (SVM) and decision trees. Feature selection for regression models is often done through introducing penalties on the coefficients. Typical penalties include ridge regression (L2 penalty) and Lasso (L1 penalty). Compared to ridge regression,

Lasso shrinks the unnecessary coefficients to zero instead of small values, which provides both intuition and good performance. Stochastic search variable selection (SSVS) uses “spike and slab” prior so that the posterior of the coefficients have some probability being 0.

Another approach is to consider the variable selection problem as model selection, i.e. put priors on models (e.g. a Bernoulli prior on each coefficient being 0) and compute the marginal posterior probably of the model given data. People then either use Markov Chain Monte Carlo to sample models from the model space and apply Bayesian model averaging, or do a stochastic search in the model space to find the posterior mode. Among non-linear models, tree models usually handles the non-linearity and variable selection simultaneously. Representative work includes decision trees, random forests, gradient boosting and Bayesian additive regression trees (BART).

3. Proposed Methodology

In this paper we have proposed following new methodologies as proposed techniques for Online Fraud Detection. They are as follows:

A) Online Equity Regression Mechanism.

B) Online Feature Selection through SSVS Mechanism.

A) Online Equity Regression Mechanism

Consider splitting of continuous time into many small equal size intervals as in [11]. For every interval we may have many expert labeled cases which indicate whether they are fraud or not. At time interval t suppose there are nt observations. Let us denote the i -th binary observation as y_{it} . If $y_{it} = 1$, the case is fraud, otherwise it is non-fraud. Let the feature set of case i at time t be \mathbf{X}_{it} . The online fraud detection model as in can be written as

$$\mathbf{P} [y_{it} = 1 | \mathbf{x}_{it}, \alpha_t] = \Phi (\mathbf{x}_{it}^T \alpha_t), \quad (1)$$

Where $\Phi (\cdot)$ is the cumulative distribution function of the standard normal distribution $N (0, 1)$, and α_t is the unknown regression coefficient vector at time t . By data augmentation the online fraud detection model can be expressed in a hierarchical form as follows:

For each observation i at time t assume a latent random variable Z_{it} . The binary response y_{it} can be viewed as an indicator of whether $z_{it} > 0$, i.e. $y_{it} = 1$ if and only if $Z_{it} > 0$. If $Z_{it} \leq 0$, then $y_{it} = 0$. Z_{it} can then be modeled by a linear regression as .

$$z_{it} \sim N (\mathbf{x}_{it}^T \alpha_t, 1), \quad (2)$$

In a Bayesian modeling framework it is common practice to put a Gaussian prior on α_t .

$$\alpha_t \sim N (\mu_t, \Sigma_t), \quad (3)$$

Where μ_t and Σ_t are prior mean and prior covariance matrix respectively.

B) Online Feature Selection through SSVS Mechanism

For regression problems with many features, proper shrinkage on the regression coefficients as in is usually required to avoid over-fitting. For instance, the two common shrinkage methods are L2 penalty (ridge regression) and L1 penalty (Lasso) as in. Experts often want to monitor the importance of rules so that if any adjustments are required they can modify it for effective use. By this expert as in [11] can add new rules or change rules. However, the fraudulent sellers change their behavioral pattern quickly: some rule-based features that do not help today might help a lot tomorrow. For that it is necessary to build an online feature selection framework and intuition as in. At time t , let α_{jt} be the j -th element of the coefficient vector α_t . Instead of putting a Gaussian prior on α_{jt} , the prior of α_{jt} now as follows.

$$a_{jt} \sim p_{0jt} 1(a_{jt} = 0) + (1 - p_{0jt}) N(\mu_{jt}, \sigma_{jt}^2), \quad (4)$$

Where p_{0jt} is the prior probability of a_{jt} being exactly 0, and with prior probability $1 - p_{0jt}$, a_{jt} is drawn from a Gaussian distribution with mean μ_{jt} and variance σ_{jt}^2 . Such prior is called the “spike and slab” as in [8] but how to embed it to online modeling has never been explored before.

4. Implementation Modules

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and its constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

The proposed consists of totally 4 Modules:

- a. Registration & Authentication Module
- b. Login Module
- c. Find Fraud Module
- d. User Complaint Module

a) Registration & Authentication Module

In this module, the user is initially registered into the web site in order to enter into his individual account for performing the operations. The user may be a seller, customer (I.e Buyer). When user comes first time for browsing, he should enter into this module in order to create new profile registration to that particular user. After profile registration is successful with all valid field and user id and password, the user will be entered into the online model based on his role either as a buyer or as a seller if he is giving a valid userid and password, if not login fails.

b) Login Module

In this module, after successful registration of all sellers and users, the administrator login into the model with his user id and password for performing various operations like view all sellers and giving authorization access for the newly registered sellers. If the authorized access is not given to sellers by admin, even they have userid and pass word, they can't able to enter into their individual accounts for creating new products and offers for the products. Hence if the authorization is given by admin to the seller then he/she can able to place new products and offers for online shopping available users who wish to do purchase in this site.

c) Find Fraud Module

In this module, if one case is labeled as fraud by human experts that are maintained by administrator, it is very likely that the seller is not trustable and may be also selling other frauds; hence all the items submitted by the same seller are labeled as fraud too. The fraudulent seller along with his/her cases will be removed from the website immediately once detected.

d) User Complaint Module

Buyers can file complaints to claim loss if they are recently deceived by fraudulent sellers. The administrator views the various type of complaints like Not Delivered, Product Mismatch, Poor Service and Product Damaged, and the percentage of various type complaints. The complaints values of a products increase some threshold value the administrator set the trustability of the product as Untrusted or banded. If the products set as banaded, the user cannot view the products in the website.

5. Conclusion

In this paper, we designed a new online model which is almost best suited for any type of online shopping sites in order to provide high security for the buyers from the fault users or intruders. We mainly designed proactive based online models for the auction fraud detection and avoidance of fraud. We show that our proposed

online proactive model framework is based on a real word online auction fraud detection data, which combines online feature selection, bounding coefficients from proficient knowledge and several instance learning and can extensively develop over baselines and the human-tuned model. This online modeling frame can be simply extended to several other applications like web spam detection, content optimization and so forth. . The adjustment of the selection bias in the online model training process is included to one direction and has proven to be very efficient for offline models.

6. References

- [1] Agarwal D, Chen B, Elango P. Spatio-temporal models for estimating click-through rate. In Proceedings of the 18th international conference on World wide web ACM 2009; 21-30.
- [2] Andrews S, Tsochantaridis I, Hofmann T. Support vector machines for multiple-instance learning. Advances In Neural Information Processing Systems 2003; 577-584.
- [3] D. AGARWAL, B. CHEN, AND P. ELANGO. SPATIO-TEMPORAL MODELS FOR ESTIMATING CLICK-THROUGH RATE. IN PROCEEDINGS OF THE 18TH INTERNATIONAL CONFERENCE ON WORLD.
- [4] D. Gregg and J. Scott. The role of reputation systems in reducing on-line auction fraud. International Journal of Electronic Commerce. 10(3):95-120, 2006.
- [5] Federal Trade Commission. Internet auctions: A guide for buyers and sellers. <http://www.ftc.gov/bcp/online/pubs/online/auctions.htm>, 2004.
- [6] D. Chau and C. Faloutsos. Fraud detection in electronic auction. In European Web Mining Forum (EWMF 2005), page 87.
- [7] A. Borodin and R. El-Yaniv. Online computation and competitive analysis, volume 53. Cambridge University Press New York, 1998.
- [8] H. Ishwaran and J.Rao. Spike and Slab variable selection: Frequentist and Bayesian strategies. The Annals of Statistics,33(2): 730-733 ,2005.
- [9] L. Breiman. Random forests. Machine learning, 45(1):5–32, 2001.
- [10] C. Chua and J. Wareham. Fighting internet auction fraud: An assessment and proposal. Computer, 37(10):31–37, 2004.
- [11] R. Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) ,58(1): 267-288, 1996.